POLYP DETECTION IN CT COLONOGRAPHY BASED ON SHAPE CHARACTERISTICS AND KULLBACK-LEIBLER DIVERGENCE

Ju Lynn Ong, Abd-Krim Seghouane

National ICT Australia*
Canberra ACT 2600, Australia
julynn.ong@nicta.com.au,
abd-krim.seghouane@nicta.com.au

Kevin Osborn

Canberra Imaging Group
77, Denison Street
Deakin, ACT 2600, Australia
kosborn@cig.com.au

ABSTRACT

As an alternative procedure to the current methods which consider only the mean values of shape features to globally characterize a candidate shape polyps, probability density functions (PDFs) of some feature variables constructed based on Gaussian and mean curvatures are used to characterize the global shape of a candidate lesion. The decision on whether or not this candidate lesion is a polyp is made by comparing the density functions of the considered shape feature variables to reference PDFs of the same variables obtained from a pre-constructed polyp/non polyp data base. The Kullback-Leibler divergence is used as a dissimilarity measure to compare these PDFs and make a decision based on closeness. Experiments carried out on real data are used to illustrate the effectiveness of the proposed method in comparison to existing ones.

Index Terms— Image Shape Analysis, Probability Measures

1. INTRODUCTION

Colorectal carcinoma or colon cancer is the second leading cause of cancer deaths in Australia, with approximately 4700 deaths each year [1].

Studies so far have shown that this type of cancer often arises from pre-existent adenomatous polyps-an abnormal growth of tissue. These polyps are believed to proceed through a genetic alteration stage in which small adenomas ($<$5mm) grow into large adenomas ($>$10mm), then into non invasive carcinoma and finally invasive carcinoma. The time taken for these small adenomas to develop into terminal stage cancer is 10-15 years on average, and a study has shown that early removal of colonic polyps resulted in a decline in mortality of about 76-90% [2].

For this reason, checking for precancerous or cancerous cells in the colon is highly recommended. Computed Tomographic (CT) Colonography is a recent, non-invasive technique that produces cross-sectional CT data scans which can be used to reconstruct volumetric colonic data of the patient for further processing and analysis. Several computer-aided diagnosis (CAD) systems have been developed to aid the detection of these adenomatous polyps in order to be removed in the early stage. However, polyp detection is a challenging process because they come in various shapes and sizes. In addition, residual materials and folds in the colonic wall may also resemble polyps and it is therefore important to use proper shape measures to detect these polyps.

In [3], [4] and [5], polyp candidates are detected by computation of 3-D geometric shape features such as curvedness and volumetric shape index based on the curvature characteristics at each voxel of the colon volume or vertex of the colon surface. These methods demonstrate acceptable sensitivity and moderate specificity by only using the mean value as a global shape characteristic. However, solely using the first or second moment to characterize a distribution is not efficient, as two Gaussian densities can have the same first moment but different second moment. It is also difficult to select a good threshold based only on a small polyp database. Therefore we predict that better global shape characteristics than the first and second moments will improve the specificity and sensitivity of polyp detection procedures. For this reason, we propose to use instead the probability density functions (PDFs) of the 3-D geometric shape features approximated by normalized histograms as an efficient global characteristic or a unique signature for a candidate polyp shape in this paper. PDFs have been used to describe the global geometric properties and signature of an object [6]. The problem is reduced to the comparison of two PDFs and is relatively simple to implement compared to the traditional shape matching methods of pose registration, parameterization, feature correspondence, template matching and model fitting. In this paper, the measure of how close the normalized histograms are is then computed using the Kullback-Leibler (K-L) divergence.

The rest of the paper is organized as follows. In the next section we review the 3-D geometric shape features used in

*National ICT Australia is funded by the Australian Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Center of Excellence Program.

this paper, describe the construction of the PDFs and propose a comparison technique. Section 3 illustrates the performance of the proposed method in polyp detection in comparison to another existing method employing only the mean values. Concluding remarks are given in Section 4.

2. METHOD

2.1. Computation of 3-D Geometric Features

3-D geometric shape features that characterize polyp and non polyp lesions first need to be selected and computed in order to identify the lesions in the later stage. In [7], the D1 PDF constructed using distances between a fixed point and one random point on the surface described in [6] has been used with good results for polyps greater than 10mm in diameter. In this paper, Gaussian and mean curvature values computed from the surface of the lesion are used instead. Surface curvature values can better measure the shape characteristics of a local patch at a point which can then be used to distinguish between the different lesions (cylindrical folds, spherical polyps, etc) than the D1 measure. While the D1 measure also requires an extra normalizing step to ensure scale invariance, the range and sign of curvature values that are used in our method can be predicted for similar shapes regardless of the lesion size.

Gaussian and mean curvature values are two important notions of describing the local curvature of a surface at a point, the former being an intrinsic property of a surface and is a measure of the spherical spread of surface normals, while the latter an extrinsic property and is a measure of the average of dihedral angles [8]. These values at each vertex are estimated via the Gauss-Bonnet scheme described in [9]. Then the computations of the Gaussian, K and mean, H curvatures at a vertex p respectively are given by

$$K(p) = \frac{2\pi - \sum_{i=0}^{n-1} \alpha_i}{\frac{1}{3}A} \tag{1}$$

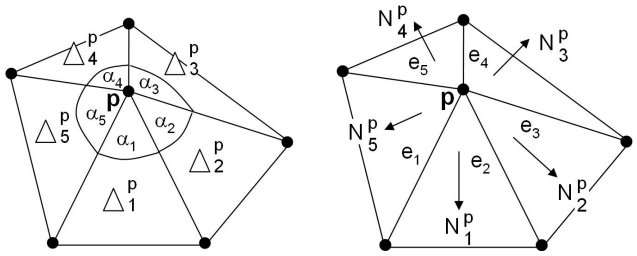
where K is assumed to be constant in the local neighborhood of a vertex p and its immediate 0 to $n-1$ neighbors, α_i are the internal angles between two successive neighboring edges and A is the accumulated areas of the neighboring triangles (Δ_i^p) as shown in Figure 1(a), and

$$H(p) = \frac{\frac{1}{4}\sum_{i=0}^{n-1} \parallel e_i \parallel \beta_i}{\frac{1}{3}A} \tag{2}$$

where $\parallel e_i \parallel$ denotes the magnitude of e_i and β_i measures deviations from the normal vectors of two neighbors using $\beta_i = \angle(N_i^p, N_{i+1}^p)$ as shown in Figure 1(b).

The two principal curvatures associated with vertex p, $k_1(p)$ and $k_2(p)$ can then be calculated using [3]

$$k_1(p) = H(p) + \sqrt{H^2(p) - K(p)} \tag{3}$$



(a) Angles α_i and areas Δ_i^p (b) Edges e_i and normals N_i^p

Fig. 1. Notations for surface curvature estimation via the Gauss-Bonnet scheme.

$$k_2(p) = H(p) - \sqrt{H^2(p) - K(p)} \tag{4}$$

Based on these principal curvature values, the shape index SI, which is a measure of the shape of the local neighborhood at a vertex, the curvedness CV, a measure of the size of that particular shape at each vertex p and the sphericity ratio SR, a measure of uniformity of the polyp shape can be computed as described in [3] and [5]

$$SI(p) = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \tag{5}$$

$$CV(p) = \sqrt{\frac{k_1(p)^2 + k_2(p)^2}{2}} \tag{6}$$

$$SR(p) = |\frac{(k_2(p) - k_1(p)}{H(p)}| \tag{7}$$

Each distinct shape class can be described by a unique value of SI. [3] describes the relationship of shapes and their corresponding shape indices, for example an SI of 1 corresponds to the cup shaped isosurface that often describe colonic polyps and an SI of 0.75 corresponds to the rut shape that often describe cylindrical folds in the colon. Perfect planes however, have an indeterminate SI due to zero-valued principal curvatures and are neglected in our computations. Since we are only interested in the detection of spherical polyps, this is considered a reasonable concession.

2.2. Colon Wall Extraction

Typical CT colonography datasets usually consists of between 200-350 axial CT scans in a 512 x 512 matrix. These are linearly interpolated and the colonic region was segmented from the colon CT volume dataset using a -700HU threshold known to correspond to the air filled regions within the colon. Next a polygonal mesh of the colon wall surface was extracted using the marching cubes algorithm and smoothed using a Gaussian filter with a radius of 1 voxel and standard deviation of 0.5 voxels. No decimation was performed and this entire mesh usually comprises between 800,000 to 1,000,000 polygonal vertices. This entire process was implemented in C++ using the Insight Toolkit (ITK) and Visualization Toolkit (VTK) libraries.

2.3. Construction of Reference PDFs

As in [5] and [10], regions that belong to the elliptic class of the peak subtype - $K(p) > 0$ and $H(p) < 0$ are detected and clusters which have more than 80 vertices satisfying this criteria are marked as candidate lesions to filter out isolated vertices. This arbitrary low value was chosen as all polyps covered a larger area than these 80 vertices and this assumption was valid in all our experiments. Next, instead of using the mean values of the 3-D features described in Section 2.1 to globally characterize the shape of a candidate polyp, normalized histograms of K, H, k_1, k_2, SI, CV and SR from >10 typical polyp/non polyp regions are used in this paper to approximate reference PDFs of these shape features over the entire candidate lesion. As the regions detected are on average only 80-500 vertices, all shape feature values of all vertices from these polyp/non polyp regions are used to construct histograms by counting how many vertex shape feature values fall into a certain bin. These values are then normalized to construct a reference PDF for that lesion-polyp/non polyp.

The highlighted regions in Figure 2 show the areas that fall into the elliptic class of the peak subtype detected by our CAD system. Note that these regions were cropped from the original colon surface for ease of view. Since the non polyp regions detected can assume a variety of forms with differing shape features-large folds, narrow folds and semi-flat regions as shown in Figure 2, we construct individual reference PDFs for each lesion type. An example of a polyp and narrow fold reference PDF set is shown in Figures 3 and 4. An example of the reference PDFs constructed using the mean curvatures only of the each region is given in Figure 5. It is noted that the narrow folds and semi-planar regions have also very distinct PDFs from the polyp regions as they are generally more noisy, have multiple peaks and an irregularity in shape, whilst it is more difficult to distinguish between the polyp and large fold regions just by looking at their PDFs.

2.4. Comparison of Shape Distributions

We ran a test of dissimilarity using the K-L divergence to measure the difference between a reference PDF P constructed in Section 2.2 to a new PDF Q constructed using shape feature values from the vertices of a new detected candidate lesion. The K-L divergence is then defined to be

$$D_{kl}(P \parallel Q) = \sum_i p(i) log \frac{p(i)}{q(i)} \quad (8)$$

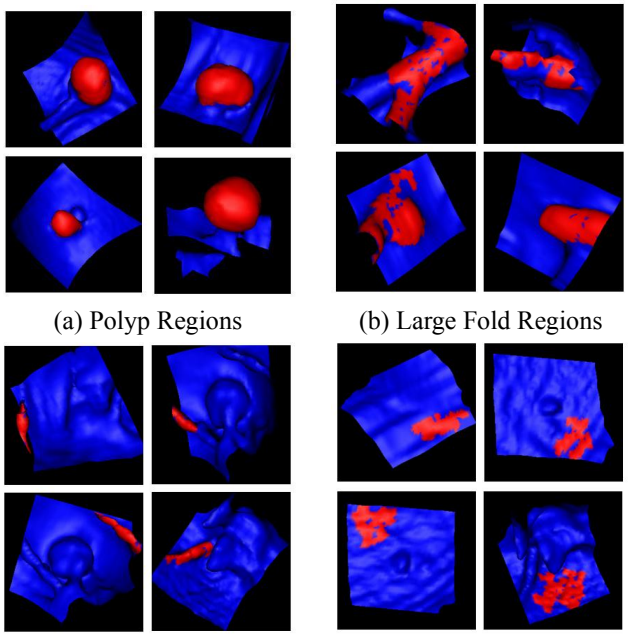
For each new candidate lesion, The K-L divergence of this PDF to the reference PDFs constructed earlier in Section 2.2 was computed and if the distance to a polyp reference PDF was smaller, the lesion was marked as a true polyp.

3. EXPERIMENTAL RESULTS

Tests were performed on 117 cropped 32x32x32 voxel colon regions obtained from actual colon datasets. Sensitivity and specificity percentages were calculated using

$$Sensitivity = \frac{TruePositives}{TruePositives + FalsePositives} \times 100 \quad (9)$$

$$Specificity = \frac{TrueNegatives}{FalsePositives + TrueNegatives} \times 100 \quad (10)$$



(a) Polyp Regions (b) Large Fold Regions

(c) Narrow Fold Regions (d) Semi-Planar Regions

Fig. 2. Polyp and non polyp regions.Candidate regions that belong to the elliptic class of the peak subtype have been highlighted.

Table 1 shows the sensitivity and specificity results obtained for each 3-D geometric shape feature computed with our method. Table 2 shows the sensitivity and specificity results obtained by taking means of the same shape features from the same database and using these as thresholds to detect polyps. Based on these results we have found that the new method employing the PDFs of H, k_2 and CV have better sensitivity compared to the method of taking mean values and shows better specificity values for all shape features.

Feature	H	k_2	CV
Sensitivity	82%	82%	89%
Specificity	72%	70%	71%

Table 1. Sensitivity and specificity using PDFs.

4. CONCLUSIONS AND FURTHER WORK

In this paper, instead of using characteristics of a set of shape features like the mean of the Gaussian curvature or the mean

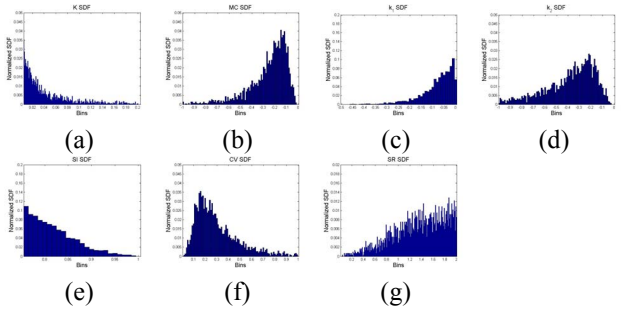


Fig. 3. PDFs constructed for our polyp database using the values of (a) K (b) H (c) k_1 (d) k_2 (e) SI (f) CV and (g) SR

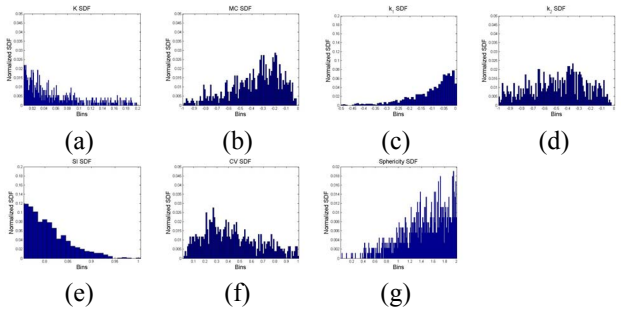


Fig. 4. PDFs constructed for a narrow fold non polyp database using the values of (a) K (b) H (c) k_1 (d) k_2 (e) SI (f) CV and (g) SR

of the shape index over a number of voxels from a specific region of the colon, namely a candidate polyp, we have used the PDFs of these shape features to globally characterize a candidate polyp. The decision on whether or not this candidate lesion is a polyp was made by comparing the PDFs using the K-L divergence of the considered shape feature values to reference PDFs of the same feature values obtained from a polyp/non polyp data base. Our experimental results show that the proposed method provides better performance for the shape features H, k_2 and CV than existing methods that only use the mean values of a set of features.

Further work also needs to be done as a measure of deviation from smoothness using the values of H, k_2 and CV to discriminate between polyps and non polyps as we have found these to be most characteristic of the shape of the lesion.

Feature	H	k_2	CV
Sensitivity	63%	53%	47%
Specificity	44%	50%	42%

Table 2. Sensitivity and specificity using mean values.

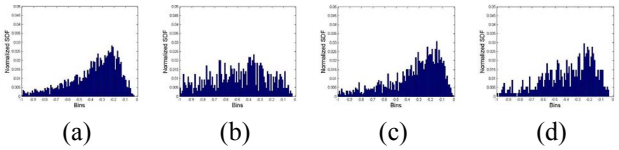


Fig. 5. PDFs constructed using mean curvature values for (a) Polyp (b) Narrow Fold (c) Large Fold (d) Semi-Planar

5. REFERENCES

[1] Australian Institute of Health and Welfare, "Cancer in Australia 2001," *www.aihw.gov.au/publications/can/ca01/ca01.pdf*

[2] C. F. Beaulieu, R. B. Jeffrey, P. Li, D. S. Paik and S. Napel, "A 3D statistical classification scheme for computer-aided detection of colonic polyps in CT images," *Stanford University, School of Medicine and School of Statistics*, 2005.

[3] H. Yoshida and J. Nappi, "Three-dimensional computer aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imag.*, vol. 20(12), pp. 1261–1274, 2001.

[4] H. Frimmel, J. Nappi and H. Yoshida, "Virtual endoscopic visualisation of the colon by shape-scale signatures," *IEEE Trans. Info. Tech.*, vol. 9(1), pp. 120–131, 2005.

[5] L. M. Pusanik, R. M. Summers and C. F. Beaulieu, "Automated polyp detector for CT colonography," *Radiology*, vol. 216, pp. 284–290, 2000.

[6] B. Chazelle, R. Osada, T. Funkhouser and D. Dobkin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21(4), pp. 807–832, 2002.

[7] O. Ghita, A. A. Miranda, T. A. Chowdhury and P. F. Whelan, "Shape filtering for false positive reduction at computed tomography colonography," *Medical Image Computing and Computer-Assisted Intervention MICCAI 2006*, pp. 84–92, 2006.

[8] J. Koenderink, *Solid Shape*, MA: MIT Press, Cambridge, 1990.

[9] O. Soldea, E. Magid and E. Rivlin, "A comparison of gaussian and mean curvature estimation methods on triangular meshes of range image data," *Comp. Vis. Image Understand.*, vol. 107, pp. 135–139, 2007.

[10] J. D. Malley, R. M. Summers and W. S. Selbie, "Polypoid lesions of airways: Early experience with computer assisted detection by using virtual bronchoscopy and surface curvature," *Radiology*, vol. 208, pp. 331–337, 1998.